

Efficient Monte Carlo characterization of quantum operations for qudits

Giulia Gualdi,^{1,2} David Licht,³ Daniel M. Reich,³ and Christiane P. Koch³

¹*Dipartimento di Fisica ed Astronomia, Università di Firenze, Via Sansone 1, 50019 Sesto Fiorentino, Italy*

²*QSTAR, Largo Enrico Fermi 2, 50125 Firenze, Italy*

³*Theoretische Physik, Universität Kassel, Heinrich-Plett-Str. 40, D-34132 Kassel, Germany*

(Dated: April 8, 2014)

For qubits, Monte Carlo estimation of the average fidelity of Clifford unitaries is efficient – it requires a number of experiments that is independent of the number n of qubits and classical computational resources that scale only polynomially in n . Here, we identify the requirements for efficient Monte Carlo estimation and the corresponding properties of the measurement operator basis when replacing two-level qubits by p -level qudits. Our analysis illuminates the intimate connection between mutually unbiased measurements and the existence of unitaries that can be characterized efficiently. It allows us to propose a ‘hierarchy’ of generalizations of the standard Pauli basis from qubits to qudits according to the associated scaling of resources required in Monte Carlo estimation of the average fidelity.

PACS numbers: 03.65.Wj, 03.67.Ac

I. INTRODUCTION

The capability to verify that a quantum operation has been properly implemented is an important building block for quantum technologies [1]. It requires evaluation of suitable performance measures such as the average fidelity or the worst case fidelity. In general, evaluating either measure scales very unfavorably in system size due to the exponential scaling of the Hilbert space dimension d with the number n of information carriers. Stochastic sampling techniques have recently allowed for impressive progress at reducing the resources required for determining the average gate fidelity for qubits [2–7]. For example, Monte Carlo estimation can be employed to determine the average n -qubit gate fidelity F_{av} [2, 3]. To this end, F_{av} is expressed either in terms of the entanglement fidelity [2, 3] or as a sum over $d(d+1)$ state fidelities in d -dimensional Hilbert space where the $d(d+1)$ states form a so-called state 2-design [8, 9]. The latter represents the optimal strategy in terms of the average number of experiments that need to be performed, the number of settings from which an experiment is drawn in the Monte Carlo procedure and the associated computational complexity [9]. The effort for estimating the average gate fidelity can be further reduced when determining bounds instead of F_{av} itself [9]. The bounds are given by two classical fidelities in Hilbert space each made up of d state fidelities [10].

These statements hold for both general unitaries and Clifford gates. However, for Clifford gates, the three approaches differ merely in the number of experimental settings; the average number of experiments is independent of system size [2, 3, 9]. As a consequence, estimating the average fidelity of a Clifford gate is a task that can be performed efficiently, i.e., with an effort that scales at most polynomially with the number of qubits.

Clifford gates represent an important subset of quantum gates – they facilitate fault-tolerant computation [11] and yield a universal set when augmented by

the proper local phasegate [12]. They can be used to prepare entangled states and perform quantum teleportation even though their computing power is not stronger than classical [13]. The striking observation that the experimental effort for Clifford gate characterization does not scale exponentially with the number of qubits is due to the property of Clifford gates to map stabilizer states into stabilizer states. This property is also exploited by another efficient method for determining the average gate fidelity, termed randomized benchmarking [6, 7].

The Clifford gate property translates, for Monte Carlo estimation of the average fidelity, into a relevance distribution which is uniform and known *a priori* [2, 3]. A uniform relevance distribution does not require sampling; and the average number of experiments becomes independent of system size. It turns out, however, that the uniformity of the relevance distribution is tied to the Pauli operators having eigenvalues ± 1 . It therefore applies to qubits but not to Hilbert spaces of prime power dimensions $d = p^n$ with p other than two. This raises the question of whether and how the Clifford property of mapping stabilizer states into stabilizer states can be exploited to efficiently estimate the average gate fidelity for qudits ($p > 2$).

Qudits in general and qutrits ($p = 3$) in particular occur naturally in many quantum systems: They can be encoded in anharmonic ladders of e.g. superconducting circuits [14, 15], in orbital angular momentum modes of photons [16, 17] or in the polarization of biphotons [18, 19]. Compared to qubits as quantum information carriers, they offer advantages in terms of increased security and higher channel capacity in quantum communication and better efficiency in quantum information, see e.g. Refs. [16–18]. Since device characterization is one of the prerequisites for any quantum information and communication architecture, it would represent a severe disadvantage of qudits if the average fidelity of qudit Clifford gates could not be determined efficiently.

Here, we demonstrate that Monte Carlo estimation of

the average fidelity can be made efficient for Clifford gates of qudits by suitable choice of the operator basis for the measurements. Based on intuition obtained for the qubit case, we show that the measurement basis needs to allow for a partitioning into $d + 1$ commuting sets of operators to ensure existence of a non-trivial class of unitaries that map stabilizer states into stabilizer states and yield a uniform relevance distribution. For qudits ($p > 2$), only unitary, non-Hermitian operators give rise to such a maximal partitioning. Two routes can be followed to obtain a practical characterization protocol from this observation: One can either construct Hermitian operators by suitable superposition of the basis unitaries; or utilize the concept of quantum circuits to simulate Hermitian measurements. We discuss both options. In general, we show that one can define a hierarchy of operator bases according to their scaling of resources in the Monte Carlo characterization of Clifford gates.

The paper is organized as follows: We start with a review of Monte Carlo estimation of the average fidelity for qubits [2, 3] in Section II. In particular, we explain the role of operator bases of Hilbert space for evaluating the relevance distribution for qubits and we show how the scaling in resources is obtained from it. We construct the operator basis for qudits in Sec. III, starting from the condition of a maximal partitioning and imposing further constraints on the operators to ensure efficient characterization for a maximal number of unitaries. We present the relevance distributions resulting from these bases and discuss the corresponding Monte Carlo procedures in Sec. IV. Section V concludes.

II. MONTE CARLO ESTIMATION OF THE AVERAGE FIDELITY FOR QUBIT CLIFFORD GATES

We first provide an overview over the general ideas underlying the Monte Carlo approach [2, 3]. Subsequently we explain, following Ref. [2], why for a Clifford gate the resources required for Monte Carlo estimation of the average fidelity do not scale exponentially with the number of qubits.

A. Recasting F_{av} in terms of measurements

We consider a system of n qubits with a Hilbert space of dimension $d = 2^n$. The associated Liouville space, of dimension d^2 , can be spanned by a complete and orthonormal operator basis W_k with $\text{Tr}[W_i W_k] = d \delta_{i,k}$ $\forall i, k = 1, \dots, d^2$. From a physical perspective, the operator basis represents the set of measurements that will have to be performed. The goal is to estimate the average fidelity F_{av} of a quantum device that is supposed to execute the gate $U \in \mathcal{U}(d)$. In other words, determining F_{av} verifies how well the actual evolution of the system, represented by the dynamical map \mathcal{D} , matches the target

U [1].

One possibility to evaluate F_{av} with a Monte Carlo procedure [2] rewrites F_{av} in terms of the entanglement fidelity F_e [20, 21],

$$F_{av} = \frac{dF_e + 1}{d + 1}. \quad (1)$$

F_e is defined as [2, 21, 22]

$$F_e = \frac{1}{d^2} \text{Tr} [\mathcal{U}^\dagger \mathcal{D}], \quad (2)$$

where \mathcal{U} denotes the unitary dynamical map corresponding to the desired gate U . A second option, using the channel-state isomorphism, interprets F_e as a state fidelity on an extended d^4 -dimensional Liouville space [3]. The two approaches are equivalent. Expanding the trace in Eq. (2) in the operator basis W_k , one obtains [2]

$$F_e = \frac{1}{d^4} \sum_{k,k'} \text{Tr}[W_k U W_{k'} U^\dagger] \text{Tr}[W_k \mathcal{D}(W_{k'})]. \quad (3)$$

The corresponding measurements are performed on inputs that have passed the device. Both are subjected to Monte Carlo sampling. Formally, the inputs are the operators $W_{k'}$. The obstacle that, in an experiment, one cannot prepare input operators is circumvented by sampling, additionally, over each input operator's eigenstates [2]. The set of inputs I consists of all $T = d^2$ (rescaled) operators $W_{k'}/d$ that constitute the orthonormal basis. In practical terms, Monte Carlo estimation of the average fidelity consists in randomly selecting pairs of input states and measurements that will be performed on the output obtained after sending the input through the quantum device. Summing up all measurement outcomes with the appropriate weights, given by the so-called relevance distribution (for details see Sec. II B below), yields the average fidelity.

The formal use of input operators, or, equivalently, the channel-state isomorphism, can be avoided by evaluating F_{av} as a state 2-design [8, 9]. Then the set of inputs I consists of $T = d(d + 1)$ regular Hilbert space states, which make up $d + 1$ mutually unbiased bases (MUB), and the average fidelity is expressed as

$$\begin{aligned} F_{av} &= \frac{1}{d(d + 1)} \sum_{j=1}^{d(d+1)} \text{Tr} [\rho_j^{ideal} \rho_j^{actual}] \\ &= \frac{1}{d^2(d + 1)} \sum_{j=1}^{d(d+1)} \sum_{k=1}^{d^2} \text{Tr} [\rho_j^{ideal} W_k] \text{Tr} [\rho_j^{actual} W_k], \end{aligned} \quad (4)$$

where $\rho_j^{ideal} = U |\Psi_j\rangle \langle \Psi_j| U^\dagger$ and $\rho_j^{actual} = \mathcal{D}(|\Psi_j\rangle \langle \Psi_j|)$. Another option is to determine bounds on the average gate fidelity instead of F_{av} itself using two classical fidelities [9, 10]. Each classical fidelity is expressed as a sum over $T = d$ input states, analogously to Eq. (4), with the states belonging to two MUB [9]. The different

sets of inputs for the three protocols result in different numbers of required experimental settings, average numbers of actual measurements, and classical computational resources [9].

B. Relevance distribution

The idea underlying the Monte Carlo approach is to treat $\text{Tr}[W_k \mathcal{D}(I_i)]$, where $I_i \in I$ denotes an element of the set of inputs, either operators or states, as a random variable. Then the average fidelity becomes the expectation value of a random variable, i.e., one can write F_{av} as

$$F_{av}^j = \sum_{i=1}^T \sum_{k=1}^{d^2} P^j(i, k) X_{i,k}, \quad (5)$$

where j indicates the specific protocol (entanglement fidelity, state 2-design, or classical fidelities). $P^j(i, k)$ is the so-called relevance (i.e., probability) distribution associated to F_{av}^j , and the $X_{i,k}$ are the values taken by the random variable X . Obviously, $\text{Tr}[W_k U I_i U^\dagger]$ will be absorbed into $P^j(i, k)$. The indices $i \in [1, T]$ and $k \in [1, d^2]$ run over the set of inputs and the set of measurements. The size of the space of Monte Carlo events, i.e., the domain of the relevance distribution, is therefore given by Td^2 . The relevance distribution $P^j(i, k)$ and random variable $X_{i,k}$ can be expressed in terms of the characteristic functions,

$$\chi_U^j(i, k) = \text{Tr}[W_k U I_i U^\dagger], \quad (6a)$$

$$\chi_{\mathcal{D}}^j(i, k) = \text{Tr}[W_k \mathcal{D}(I_i)], \quad (6b)$$

that represent the expectation value of the k th measurement after the i th input has passed the device. This allows to write

$$X_{ik} = \frac{\chi_{\mathcal{D}}^j(i, k)}{\chi_U^j(i, k)}, \quad (7a)$$

$$P^j(i, k) = \frac{1}{\mathcal{N}} \left[\chi_U^j(i, k) \right]^2 \quad (7b)$$

with \mathcal{N} ensuring proper normalization: $\mathcal{N} = d^2$ for the protocols based on the entanglement fidelity and on two classical fidelities, whereas $\mathcal{N} = d^2(d+1)$ for the protocol employing a state 2-design.

When evaluating F_{av}^j as expectation value of the random variable X taking values $X_{i,k}$ with known probability $P^j(i, k)$, one is faced with the problem that the $X_{i,k}$ cannot be accessed directly. As can be seen from Eq. (7a), they depend on another random variable, the expectation value $\text{Tr}[W_k \mathcal{D}(I_i)]$ of W_k . Due to the statistical nature of quantum measurements as well as random errors in the experiment, it will be necessary to make repeated measurements to determine $X_{i,k}$. We assume for a moment that the $X_{i,k}$ have been determined with sufficient accuracy and explain below what this assumption

entails. Provided the $X_{i,k}$ are known, Monte Carlo sampling estimates the expectation value F_{av}^j of the random variable X by a finite number of realizations,

$$F_{av}^j = \lim_{L \rightarrow \infty} F_L \quad \text{with} \quad F_L = \frac{1}{L} \sum_{l=1}^L X_{\kappa_l}. \quad (8)$$

Here, κ_l is the index corresponding to the l th input-output pair, i.e., $\kappa_l = (i_l, k_l)$. It can take on Td^2 values. The sample size L is chosen to guarantee that the probability for F_L to differ from F_{av}^j by more than ϵ is less than δ . The key point of the Monte Carlo approach is that while the size of the event space scales with the system size d , L depends only on the desired accuracy ϵ and confidence level δ and is independent of d .

The number of actual experiments that will have to be performed on average, will, however, depend on the system size, i.e., scale exponentially with the number of qubits, for general unitaries. This is due to the X_{κ_l} being known only approximately and can be seen as follows: The finite accuracy of the X_{κ_l} gives rise to an approximation of F_L , $\tilde{F}_L = \frac{1}{L} \sum_{l=1}^L \tilde{X}_{\kappa_l}$, where the tilde indicates approximate values. Therefore, in addition to ensuring that F_L approximates F_{av}^j with an error of at most ϵ , one also must guarantee that \tilde{F}_L approximates F_L with the desired accuracy. This implies repeated measurements for a given element κ_l ($l = 1, \dots, L$) of the Monte Carlo sample. Denoting the number of respective measurements by N_l , the total number of experiments is given by $N_{exp} = \sum_{l=1}^L N_l$. It can be shown [2, 3] that choosing

$$N_l = \frac{1}{\epsilon L [\chi_U^j(\kappa_l)]^2} \log \left(\frac{4}{\delta} \right) \quad (9)$$

guarantees the approximations of F_L by \tilde{F}_L and of F_{av}^j by F_L to hold with the desired confidence level.

Since Monte Carlo estimation is carried out by randomly drawing L times an event from the Td^2 -dimensional space of events, sampling requires \mathcal{C}_{sample} classical computational resources. The sampling step is, for a general unitary, not efficient since the dimension d of the state space scales exponentially in the number of qubits. Indeed, computing $\chi_U^j(i, k)$ requires to manipulate exponentially large matrices an exponential number of times. Note that while the sampling procedure will select only some of the settings, the ability to implement all of them is nevertheless required. The total number of measurements $\langle N_{exp} \rangle$ that needs to be carried out on average is given by summing over N_l which in turn is inversely proportional to the weight of the setting in the relevance distribution, cf. Eq. (9). The scaling of resources required to estimate the average fidelity is therefore strictly connected to the specific features of the relevance distribution, or more specifically, of the characteristic function $\chi_U^j(i, k)$ of the target unitary U in the chosen measurement basis W_k . If that basis allows many

$\chi_U^j(i, k)$ to vanish and those that do not vanish to decrease at most polynomially with the number of qubits, then the estimation procedure is efficient.

C. The special case of Clifford qubit gates

Clifford gates acting on n qubits are special in that they yield a relevance distribution which has many zeros and all non-zero values are identical. This in turn implies that the characterization of Clifford operations is efficient, i.e., the average number of experiments is independent of the number of qubits n and the classical computational effort scales only polynomially in n . In order to see why this is the case we briefly review the definitions of the Pauli group and the Clifford group as well as the action of the Clifford group on Pauli measurements and their eigenstates. Pauli observables, i.e., tensor products of single-qubit Pauli operators, represent the natural measurements in the logical basis and thus constitute the standard measurement basis for n qubits. This measurement basis can be considered 'minimal' in the sense that it only assumes the ability of implementing single-qubit gates and readout with no need for entangling operations [35].

The set of Pauli measurements $\bar{\mathcal{P}}$ acting on n qubits is defined as $\bar{\mathcal{P}} = \{\bar{P}_i = \bigotimes_{k=1}^n \sigma_{i_k}\}_{i=1}^{d^2}$ where each σ_{i_k} represents a single-qubit Pauli operator acting on the k th qubit, i.e., $i_k \in \{0, x, y, z\}$. The operators in $\bar{\mathcal{P}}$ generate the Pauli group $\mathcal{P} = \{P_k = i^a \omega^b \bar{P}_j; 0 < k \leq 4d^2\}$ with $a, b = 0, 1, j = 1, \dots, d^2$, $\omega = \exp(i\pi)$ and matrix multiplication being the group operation. It is useful to introduce sets \mathcal{W}_A of d pairwise commuting Pauli measurements. For example, \mathcal{W}_z comprises the d different tensor products made up of identities and σ_z 's.

The action of any transformation U_C belonging to the Clifford group is to map an element P_i of \mathcal{P} into another element P_k of \mathcal{P} . In other words, the Clifford group is the normalizer $\mathcal{N}(\mathcal{P})$ of the Pauli group in $U(d)$ since it leaves \mathcal{P} invariant under conjugation. This implies for the orthonormal basis of Pauli measurements $\bar{\mathcal{P}}$ that each element of $\bar{\mathcal{P}}$ is mapped into another element from this set up to a phase factor, i.e., up to a permutation of eigenvalues [23],

$$U_C \bar{P}_k U_C^\dagger = \omega^a \bar{P}_i; \quad a = 0, 1. \quad (10)$$

Clifford operations can also be defined in terms of their action on stabilizer states, i.e., in terms of their action on the joint eigenbasis of a set \mathcal{W}_A [3, 23]. One needs to fix a particular eigenbasis because each Pauli measurement acting on more than one qubit is degenerate; and it is thus not possible to characterize the action of a Clifford operation on a generic eigenbasis of a generic Pauli operator. Indeed, a Clifford operation maps joint eigenstates of the set \mathcal{W}_A into joint eigenstates of the set $\mathcal{W}_{A'}$, with either $A = A'$ or $A \neq A'$ [23, 24]. In general, one can partition the set of Pauli measurements $\bar{\mathcal{P}}$ into $d+1$ commuting sets \mathcal{W}_A , i.e., $\bar{\mathcal{P}}$ exhibits the so-called maximally

partitioning property [25]. Each partitioning defines a unique choice of $d+1$ joint eigenbases which are mutually unbiased with respect to each other [25, 26]. The maximally partitioning property ensures that, if a state $|\psi_i^A\rangle$ is a joint eigenvector of the operators in \mathcal{W}_A , its expectation value vanishes for all Pauli measurements outside of \mathcal{W}_A [36]. This can be seen as follows: If the operator basis is maximally partitioning, all operators outside of \mathcal{W}_A can be expressed in terms of an eigenbasis which is mutually unbiased with respect to $\{|\psi_i^A\rangle\}$. We recall that two complete and orthonormal bases A, A' on a d -dimensional Hilbert space are mutually unbiased if and only if

$$|\langle \psi_i^A | \psi_j^{A'} \rangle| = 1/\sqrt{d} \quad (11)$$

for all $|\psi_i^A\rangle \in A$, $|\psi_j^{A'}\rangle \in A'$ [27]. For a generic Pauli measurement belonging to the commuting set $\mathcal{W}_{A'}$, $\bar{P}_k = \sum_l \lambda_l^k |\psi_l^{A'}\rangle \langle \psi_l^{A'}|$, the expectation value is given by

$$\text{Tr} [\bar{P}_k |\psi_i^A\rangle \langle \psi_i^A|] = \sum_{j,l=1}^d \lambda_l^k |\langle \psi_i^A | \psi_l^{A'} \rangle|^2.$$

If $\mathcal{W}_A \neq \mathcal{W}_B$, then $|\langle \psi_i^A | \psi_j^{A'} \rangle|^2 = 1/d$ and

$$\text{Tr} [\bar{P}_k |\psi_i^A\rangle \langle \psi_i^A|] = \frac{1}{d} \sum_{l=1}^d \lambda_l^k = 0$$

since Pauli measurements are traceless. Therefore

$$\text{Tr} [\bar{P}_k |\psi_i^A\rangle \langle \psi_i^A|] = \begin{cases} \omega^a & \text{if } \bar{P}_k \in \mathcal{W}_A \\ 0 & \text{otherwise} \end{cases}. \quad (12)$$

Equation (12) is a consequence of the fact that measurements associated to MUB span orthogonal subspaces [27].

In the context of Monte Carlo estimation of the average gate fidelity for a Clifford gate, Eq. (12) gives rise to a uniform relevance distribution. In order to elucidate this, we distinguish whether the set of inputs I is made up of states (belonging to MUB) [9] or operators [2, 3]. In the former case, applying Eq. (12) to each state $|\psi_i^A\rangle \langle \psi_i^A| \in I$ yields for the characteristic function, cf. Eq. (6a),

$$\begin{aligned} \chi_{U_C}^j(i, k) &= \text{Tr} [\bar{P}_k U_C |\psi_i^A\rangle \langle \psi_i^A| U_C^\dagger] = \text{Tr} [\bar{P}_k |\psi_m^{A'}\rangle \langle \psi_m^{A'}|] \\ &= \begin{cases} \omega^a & \text{if } \bar{P}_k \in \mathcal{W}_{A'} \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \quad (13)$$

where $|\psi_m^{A'}\rangle$ is the m th element of the joint eigenbasis of the commuting set $\mathcal{W}_{A'}$. Inserting this into Eq. (7b) leads to

$$P_{U_C}^j(i, k) = \frac{1}{\mathcal{N}}, \quad (14)$$

i.e., the relevance distribution is uniform. It contains $\mathcal{N} = Td$ non-zero elements since for each of the T input states there are only d non-vanishing measurements.

Sampling then simply amounts to randomly drawing an index $i \in [1, T]$ to select the input state and, after calculating the output state from the action of the Clifford operation on the input state, to randomly draw an index $k \in [1, d]$ to select the output measurement from the commuting set corresponding to the output state. Uniformity of the relevance distribution implies that the sampling is independent of system size such that $\mathcal{C}_{\text{sampl}} \propto \mathcal{O}(1)$. Following the Gottesman-Knill theorem for Clifford circuits [13], the overall classical computational resources to calculate the output state scale polynomially in n .

If the set I of inputs is made up of operators, one can directly use the definition of the Clifford group as the normalizer of the Pauli group, Eq. (10), to obtain

$$\begin{aligned} \chi_{\mathcal{U}_C}^j(i, k) &= \frac{1}{d} \text{Tr} [\bar{P}_k \mathcal{U}_C (\bar{P}_i)] = \frac{1}{d} \text{Tr} [\bar{P}_k U_C \bar{P}_i U_C^\dagger] \\ &= \frac{\omega^a}{d} \text{Tr} [\bar{P}_k P_{k'}] = \pm \delta_{kk'} . \end{aligned} \quad (15)$$

Together with Eq. (7b), this leads to

$$P_{\mathcal{U}_C}^j(i, k) = \frac{1}{\mathcal{N}} \quad (16)$$

with $\mathcal{N} = d^2$. For each input operator there is only one output which leads to a non-zero outcome. Sampling amounts to randomly drawing an index $k \in [1, d^2]$ and finding i such that $\pm \bar{P}_i = U_C \bar{P}_k U_C^\dagger$. The latter can be done efficiently on a classical computer due to the Gottesman-Knill theorem [13]. Once the pair of input operator/output measurement has been selected, a second sampling step is required to randomly draw an eigenstate of the input operator \bar{P}_k . This step is computationally efficient since the spectrum of each operator corresponds to a uniform distribution. As a result, the sampling complexity $\mathcal{C}_{\text{sampl}}$ is independent of system size and the classical computational resources scale polynomially in n also for input operators [3].

The number of non-zero elements of the relevance distribution for a Clifford gate is either $Td = \mathcal{N}$, for the protocols based on input states, or $d^2 = \mathcal{N}$ for the entanglement fidelity protocol, as opposed to Td^2 for a generic unitary, independent of the protocol. This implies efficient scaling of the average number of experiments $\langle N_{\text{exp}} \rangle$ that have to be carried out for Clifford gates. In general, $\langle N_{\text{exp}} \rangle$ can be estimated by averaging over the number N_l of repetitions for each setting with the weights in the averaging given by the probability distribution $P^j(i_l, k_l)$ [2, 3, 9]. For a generic unitary, this yields

$$\begin{aligned} \langle N_{\text{exp}} \rangle &= \sum_{i_l=1}^T \sum_{k_l=1}^{d^2} P^j(i_l, k_l) N_l \\ &= \frac{1}{\mathcal{N}} \sum_{i_l=1}^T \sum_{k_l=1}^{d^2} [\chi^j(i_l, k_l)]^2 \frac{4}{[\chi^j(i_l, k_l)]^2 L \epsilon^2} \log \left(\frac{2}{\delta} \right) \\ &\propto \frac{1}{\mathcal{N}} T d^2 = \begin{cases} \mathcal{O}(d^2) & \text{for operator inputs} \\ \mathcal{O}(d) & \text{for state inputs} \end{cases} . \end{aligned} \quad (17)$$

The scaling is obtained from observing that $\kappa_l = (i_l, k_l)$ can take Td^2 values whereas $\mathcal{N} = d^2$ for operator inputs and $\mathcal{N} = Td$ for state inputs and $T = d^2$ for operator inputs. For Clifford gates, due to Eq. (13), respectively, Eq. (15), this reduces to

$$\langle N_{\text{exp}} \rangle \propto \frac{1}{\mathcal{N}} \mathcal{N} = \mathcal{O}(1) . \quad (18)$$

The fact that the number of experiments that need to be carried out is independent of system size implies that estimating the average gate fidelity is maximally efficient.

III. OPERATOR BASES FOR QUDITS

The discussion in the previous section suggests that the existence of a class of unitaries for which F_{av} can be estimated with maximal efficiency is due to two fundamental ingredients: (i) existence of a non-trivial class of unitaries ($\mathbb{U}_C = \{U_j \neq \mathbb{1}\}$) which map the operator basis into itself, up to a phase-factor; (ii) uniformity of the associated relevance distribution. Condition (i) implies that the relevance distribution associated to this class of unitaries contains a reduced number \mathcal{N} of non-zero elements which leads to $\langle N_{\text{exp}} \rangle \propto \mathcal{O}(1)$. Condition (ii) ensures that also the sampling step is efficient since the coefficients of the relevance distribution are known *a priori* with no need of explicit calculation. Both these features are intimately related to the properties of the Pauli measurement basis.

Specifically, they are connected to the fact that the set of the standard Pauli measurements can be partitioned into $d + 1$ commuting sets. This can be seen as follows: As shown in the previous section, condition (ii) follows from Eq. (12) which in turn results from the standard Pauli measurements being associated to MUB that span orthogonal subspaces, i.e., from the Pauli measurements allowing for a maximal partitioning. It seems highly likely that the maximally partitioning property is also a necessary condition for (i), i.e., the existence of target unitaries which map the measurement basis into itself, up to a phase factor. The close connection between the maximally partitioning property and the existence of \mathbb{U}_C can be inferred from the fact that Clifford operations can be defined as those unitaries that map stabilizer states into stabilizer states. That is, ensuring the existence of \mathbb{U}_C corresponds to ensuring the existence of generalized stabilizer states. These are the common eigenstates of d pairwise commuting measurement operators that have a non-vanishing expectation value only on this set of operators. In other words, the generalized stabilizer states are mutually unbiased joint eigenstates. The maximally partitioning property by itself is, however, not sufficient to ensure efficient characterization. Additionally, the spectra of the measurement operators must obey certain constraints. The dependence of the relevance distribution on the spectrum of the basis operators is apparent from Eqs. (13) and (15).

In order to determine whether there exist qudit operations that can be efficiently characterized, we thus need to find a suitable generalization of the Pauli measurements. Since Clifford gates are defined in terms of the measurement basis, cf. Eq. (10), this implies also identification of the class of unitaries \mathbb{U}_C that corresponds to the specific choice of measurement basis. Unfortunately, it is not possible to generalize all properties of the standard Pauli measurements for qubits to higher dimensions. Most notably, for $d > 2$ and $d \neq 2^m$ with m a positive integer, unitarity and Hermiticity cannot be enforced at the same time on an orthonormal and complete operator basis. Hence, when replacing qubits by qudits, it is crucial to understand what are the properties of the standard Pauli measurements that the generalized operator basis must retain for efficient estimation of the average fidelity. Moreover, it is important to determine how different features of the operator basis affect the scaling of resources of the Monte Carlo procedure. For the latter, we distinguish between the average number $\langle N_{exp} \rangle$ of experiments and the classical computational resources \mathcal{C}_{sampl} needed for the sampling. $\langle N_{exp} \rangle$ becomes independent of system size if the relevance distribution has the minimal number of non-zero elements, cf. Eq. (18). Efficient sampling in the standard MC approach requires in addition that the relevance distribution is uniform.

To identify the generalized measurement operators and the associated unitaries that leave it invariant under conjugation, we start from what we argue to be the fundamental requirement for efficient characterization – existence of $d+1$ MUB. Since they are the joint eigenbases of the measurement operators in the commuting sets of the maximally partitioning basis, the unitaries that map the operator basis onto itself should also map the set of $d+1$ MUB into itself. We utilize this property to determine candidates for the class of unitaries that can be characterized efficiently in Sec. III A. In particular, we show that any change of basis between two bases in the set of MUB leaves this set invariant. In Sec. III B we discuss the construction of an operator basis out of the $d+1$ MUB and the difficulty of guaranteeing the maximal partitioning property for the operator basis. We therefore distinguish between the single qudit and multiple qudit cases and impose the conditions for $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ at the single qudit level in Sec. III C. This ensures the average number of experiments to be independent of system size for those $U_j \in \mathbb{U}_C$ that are given by tensor products of single qudit unitaries. The conditions allow for both unitary and Hermitian operator bases. In order for \mathbb{U}_C to also comprise entangling operations, we need to impose the conditions for $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ at level of multiple qudits in Sec. III D. These conditions also allow for both unitary and Hermitian operator bases. However, it is not clear if a Hermitian basis satisfying these constraints will correspond to local measurements, whereby we mean those measurements that can be expressed as tensor products of single-qudit operators. Most likely this is not the case.

We continue with the conditions for efficient sampling

in Sec. III E and show that in order to ensure a uniform relevance distribution, the spectrum of the measurement operators must be made up of roots of unity and zero. This together with the requirement for the operator basis to be orthonormal and traceless rules out Hermitian operators. In contrast, a unitary operator basis not only allows for efficient sampling but also maximizes the set \mathbb{U}_C and can be constructed in terms of local measurements. Clearly, the notion of unitary, non-Hermitian measurements is non-standard. We therefore discuss the experimental implementation of such measurements in Sec. III F.

A. Unitary transformations between two MUB

We denote the set of $d+1$ MUB by \mathcal{M} . Since, on a d -dimensional Hilbert space, $d+1$ MUB are guaranteed to exist only if d is equal to a prime number or a power of a prime number [27], we restrict our investigation to p -level systems with p prime (qupits). We examine the properties of unitary transformations that map two bases in \mathcal{M} into each other. In particular, any such transformation is a mapping from \mathcal{M} into itself. Or more formally:

Proposition 1: Consider a basis $A_j \in \mathcal{M}$, $j \in [1, d+1]$, with elements $|\psi_k^j\rangle$, $k \in [1, d]$. Any unitary transformation between the elements of A_j and $A_{j'} \in \mathcal{M}$,

$$U_{jj'} = \sum_{k=1}^d |\psi_k^{j'}\rangle \langle \psi_k^j|, \quad (19)$$

will also be a unitary transformation between the elements of $A_i \in \mathcal{M}$ and $A_{i'} \in \mathcal{M}$ with $i' = i + (j - j')$ for each $i \in [1, d+1]$ and the sum modulo $d+1$ i.e.,

$$U_{jj'} = U_\delta, \quad (20)$$

with $\delta = j' - j$.

We prove Proposition 1 in Appendix A 1 and provide here an intuitive interpretation. Visualizing the indices j of the bases in \mathcal{M} geometrically as points on a line, mutual unbiasedness implies that all points are equally spaced and therefore must lie on a circle. The freedom in the phase factor of the overlap between two elements of two MUB, cf. Eq. (11), accounts for the number of steps separating the points on the circle. Given such a regular structure, any transformation which maps basis $A_j \in \mathcal{M}$ into basis $A_{j'} \in \mathcal{M}$ can be interpreted as a shift of $j' - j$ steps on the circle, regardless of the starting point. Hence, any shift of δ steps on the circle corresponds to a mapping, modulo $d+1$, between any two bases in \mathcal{M} whose corresponding indices i, i' are δ steps apart.

We show in Appendix A 2 that the unitaries defined by Eq. (20) can be decomposed into a transformation U_δ^0 , which maps the k th element of basis A_i into the k th element of $A_{i+\delta}$, and a permutation of the elements of the two bases. We then prove in Appendix A 2 that the unitaries defined by Eq. (20) form a group under matrix multiplication, $\mathbb{U}_\delta^{\Pi} = \{U_\delta\}$. It can be interpreted as

the composition of the group of permutations with the group of transformations $\mathbb{U}_\delta^0 = \{U_\delta^0\}$. The unitaries in \mathbb{U}_δ^Π are the candidates for \mathbb{U}_C , hence for efficient characterization, once an operator basis is constructed from the MUB.

B. Maximally partitioning operator basis

Given a set \mathcal{M} of $d + 1$ MUB, an operator basis can be constructed in terms of projectors onto the states of the MUB. This operator basis is, by construction, maximally partitioning. We recall the formal definition of a maximally partitioning operator basis [25, 26]:

Definition: An orthonormal and complete operator basis \mathbb{B} on a d -dimensional Hilbert space is maximally partitioning if there exists a $d + 1$ -dimensional set $\mathcal{M} = \{A_j\}_{j=1}^{d+1}$ of mutually unbiased bases $A_j = \{|\psi_k^j\rangle\}_{k=1}^d$ such that every operator in \mathbb{B} can be expressed as

$$B_i^j = \sum_{k=1}^d \lambda_{i,k}^j |\psi_k^j\rangle \langle \psi_k^j|. \quad (21)$$

In Eq. (21), λ^j is a $d \times d$ matrix whose rows are orthogonal. Each entry $\lambda_{i,k}^j$ corresponds to the k th eigenvalue of the i th operator in \mathbb{B} sharing the eigenbasis $\{|\psi^j\rangle\}$, i.e., belonging to the commuting set \mathcal{W}_j . In particular, since the first row of each λ^j corresponds to the spectrum of the identity, $\sum_{k=1}^d \lambda_{i,k}^j = 0$ for each $j \in [1, d + 1]$ and $i \in [2, d]$.

The identity needs to be included in the operator basis since it is left invariant by any unitary transformation and is diagonal in each of the bases in \mathcal{M} . Orthogonality of the rows of λ^j guarantees orthonormality of the operators within the same commuting set. The condition $\sum_{k=1}^d \lambda_{i,k}^j = 0$ ensures that all operators are orthogonal to the identity as well as that operators in different commuting sets are orthogonal.

In practical device characterization, the measurement operators should be tensor products of single-qubit operators. Then the measurements are local in the sense that each operator can be measured in a separable eigenbasis. The construction of an operator basis from the MUB which obeys the tensor product structure is far from straightforward. The proof of Ref. [27] ensures existence of the set of MUB but does not provide a prescription on how to actually construct the corresponding observables. For unitary operators, such a prescription is found in Ref. [26] starting from a maximally partitioning basis for a single qubit: It can be shown that the maximally partitioning property is preserved under the tensor product by making explicit use of unitarity of the single-qubit operator basis. The maximally partitioning basis for multiple qubits is then obtained by tensor products of the single-qubit unitary basis operators [26, 28]. A weaker version of this result holds also for other maximally partitioning bases, for example Hermitian ones:

Given the spectral decomposition (21), the λ^j matrices for multiple qubits can be constructed as tensor products of the λ^j matrices for $n = 1$ since orthonormality and completeness of the operator basis are preserved under tensor product. However, this does not ensure that the maximally partitioning operator basis itself can be constructed as tensor products of the single-qubit operators. In general, that is, without making any assumption on the spectra of basis operators, one obtains only $p + 1$ out of the $p^n + 1$ bases in \mathcal{M} by tensor products. This is not enough to ensure a maximal partitioning for the resulting operator basis. While it seems reasonable to expect that the maximally partitioning property is preserved only for unitary operators, it remains an open question whether this holds also for an Hermitian operator bases and if so, under which spectral conditions.

We therefore distinguish between imposing the maximally partitioning property at the single at the multi-qubit level. If only the single-qubit operator basis needs to give rise to a maximal partitioning, the multi-qubit operator basis which is constructed by tensor products from the single-qubit basis is not guaranteed to inherit this property. This implies that only unitaries that are themselves tensor products, i.e., non-entangling operations, yield $\langle N_{exp} \rangle \propto \mathcal{O}(1)$.

C. Ensuring $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ at the single qubit level

The average number of experiments required to characterize a unitary transformation, $\langle N_{exp} \rangle$, becomes independent of system size if the relevance distribution has a reduced number, \mathcal{N} , of non-zero elements. We now determine the corresponding conditions on the operator basis \mathbb{B} . To differentiate between single and multiple qubits, we indicate the dependence of the operator basis on the number n of qubits by $\mathbb{B}(n) = \{B_i(n)\}_{i=1}^{d^2}$ where $d = p^n$, $n \geq 1$. Analogously for the group of unitaries that leaves $\mathbb{B}(n)$ invariant. The conditions at the single qubit level are given by

Theorem 1: For any n , a non-trivial class of unitaries, $\mathbb{U}_C(n)$, for which $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ exists if

1. the operator basis for a single qubit, $\mathbb{B}(1)$, is maximally partitioning,
2. all single-qubit λ^j 's in the decomposition (21) are equal.

We prove this theorem in Appendix A 3. The idea underlying the proof is the following: Conditions 1 and 2 ensure that the single-qubit operator basis $\mathbb{B}(1)$ is left invariant by the group of transformations $\mathbb{U}_\delta^0(1)$. Consider the multiple-qubit operator basis $\mathbb{B}(n)$ that is obtained from tensor products of the elements of $\mathbb{B}(1)$. By construction it is left invariant by the set of transformations $\tilde{\mathbb{U}}_\delta^0(n)$, obtained from tensor products of the elements of $\mathbb{U}_\delta^0(1)$. The transformations in $\tilde{\mathbb{U}}_\delta^0(n)$ obey a relation analogous to Eq. (15) and thus yield an average number

of experiments that is independent of system size for the protocol based on the entanglement fidelity.

By construction, the operators in $\mathbb{B}(n)$ admit the existence of a set $\mathcal{M}^{sep}(n)$ of $p + 1$ separable mutually unbiased joint eigenbases. These are obtained from tensor products of the elements of the single-qubit set of MUB, $\mathcal{M}(1)$. The set $\mathcal{M}^{sep}(n)$ is mapped into itself by the transformations in $\tilde{\mathbb{U}}_\delta^0(n)$, and the states belonging to it obey a relation analogous to Eq. (12). Hence, if the characterization protocol does not require more than $p + 1$ MUBs, the relevance distribution of the transformations in $\tilde{\mathbb{U}}_\delta^0(n)$ has \mathcal{N} non-zero elements and $\langle N_{exp} \rangle \propto \mathcal{O}(1)$. This is the case for the protocol based on the two classical fidelities but not for the two-design protocol. Since the latter requires $d + 1$ MUB, it can not be used in combination with an operator basis that only ensures existence of $p + 1$ MUB.

Conditions 1 and 2 thus ensure the existence of a group of unitaries, $\tilde{\mathbb{U}}_\delta^0(n)$, that lead to $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ for the protocol sampling over the eigenstates of input operators [2, 3] and for the protocol based on the two classical fidelities [9]. The group of transformations $\tilde{\mathbb{U}}_\delta^0(n)$ represents only a subgroup of the group $\mathbb{U}_\delta^0(n)$ since the latter must also contain entangling operations, i.e., operations which cannot be obtained as tensor products of single-qubit unitaries. This follows from the proof of Ref. [28] showing that, for $n > 1$, bases with a different entanglement structure coexist within the same set of MUB $\mathcal{M}(n)$. Therefore the group $\mathbb{U}_\delta^0(n)$ includes transformations mapping two bases with a different entanglement content into each other, i.e., entangling operations.

Theorem 1 is compatible with both real and complex spectra of the measurement operators, i.e., with both unitary and Hermitian operator bases. However, for qutrits ($p = 3$), the Gell-Mann basis, i.e., the basis of the standard generators of $SU(3)$, does not fulfill the conditions of Theorem 1 since the eigenvectors of the Gell-Mann operators are not mutually unbiased. This also implies that such a basis cannot be used with the two protocols based on input states [9] for any unitary.

The conditions in Theorem 1 define a minimal underlying regular structure of the operator basis. We assume that in absence of such a regular structure it is not possible to find any transformation, besides the identity, that maps the full operator basis into itself. Conditions 1 and 2 then endow a generic operator basis with the most general regular structure it can have that allows for a relevance distribution with a reduced number of elements at least in some protocol, at least for some unitaries.

D. Ensuring $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ at the level of n qubits

In order to achieve a number of experiments that is independent of system size for any number of qudits and independent of the protocol, the operator basis needs to allow for a maximal partitioning for every n . This is expressed by

Theorem 2: A non-trivial class of unitaries, \mathbb{U}_C , for which the scaling of $\langle N_{exp} \rangle$ is $\mathcal{O}(1)$, independent of the characterization protocol, exists if

1. the operator basis $\mathbb{B}(n)$ is maximally partitioning,
2. all λ^j 's in the decomposition (21) are equal.

This class of unitaries includes entangling operations.

This theorem can be proven in exactly the same way as for single qubits in Appendix A 3, i.e., by substituting the single-qubit operators in Eq. (A12) by multi-qubit operators. Assuming the operator basis to be maximally partitioning for every n ensures that one can construct the full set $\mathcal{M}(n)$ of MUB out of the joint eigenbases of the operators in $\mathbb{B}(n)$. This implies that $\mathbb{U}_\delta^0(n) \subseteq \mathbb{U}_C(n)$ for all protocols. The set $\mathbb{U}_\delta^0(n)$ includes entangling operations since the MUB in $\mathcal{M}(n)$ have different entanglement content [28].

Theorem 2 is compatible with both real and complex spectra of the measurement operators. However, it might not be possible to obtain a Hermitian basis from tensor products of the single-qubit bases which allows for a maximal partitioning. In that case, the Hermitian operators would not correspond to local measurements. For an operator basis that gives rise to a maximal partition and is constructed in terms of tensor products of single-qubit operators, condition 2 of Theorem 2 translates into the requirement that all single-qubit operators have the same spectrum.

So far we have identified a set of conditions that guarantee the average number of experiments in Monte Carlo estimation of F_{avg} to be independent of system size, $\langle N_{exp} \rangle \propto \mathcal{O}(1)$, for certain unitaries. This corresponds to a first step towards efficient Monte Carlo characterization. The additional condition of a uniform relevance distribution, that ensures the classical computational resources to scale at most polynomially in n , requires additional constraints on the spectra of the measurement operators.

E. Ensuring efficient sampling: The uniform relevance distribution

Efficient sampling requires a uniform relevance distribution which together with tracelessness and orthonormality of the operator basis, implies the measurement operators to have the same spectrum, up to a phase factor, with the modulus square of each eigenvalue being equal to 1. For Hermitian operators, uniformity of the relevance distribution combined with the constraint of tracelessness, implies that the spectrum of each basis operator must be the same and made up of an equal number of $+1$ and -1 and at least one zero. However, for $p > 2$, such a spectrum is incompatible with orthonormality of the operator basis. It is easy to check that already for a single qutrit ($p = 3$), this choice of eigenvalues does not allow to construct a $(p \times p)$ -matrix λ with orthogonal rows. This holds also for prime numbers $p > 3$. As

a consequence, enforcing the operator basis to be Hermitian rules out the possibility of obtaining a uniform relevance distribution and thus efficient sampling for any target unitary (except identity).

In contrast, for unitary measurement bases, tracelessness and unitarity imply that the spectrum of each single-qutrit operator is p -nary, i.e., made up of the p distinct p th roots of unity. Consequently the spectra of all multi-qutrit measurement operators are identical since each p th root of the identity simply appears with multiplicity p^{n-1} . Such a spectrum is also compatible with orthonormality. Indeed, using p distinct p th roots of unity, one can construct, for each of the $p + 1$ bases in $\mathcal{M}(1)$, a set of exactly $p - 1$ pairwise orthogonal traceless operators, i.e., a maximally partitioning single-qutrit basis [29]. Since the maximal partitioning is preserved under tensor product [26], a p -nary spectrum is also compatible with a multiple-qutrit operator basis that gives rise to a maximal partitioning. As a consequence, a maximally partitioning unitary basis is compatible with a uniform relevance distribution. It requires, however, a generalization of the relevance distribution given in Eq. (7b) to include complex expectation values,

$$P^j(i, k) = \frac{1}{\mathcal{N}} \left| \chi_U^j(i, k) \right|^2; \quad \chi_U^j(i, k) \in \mathbb{C}. \quad (22)$$

More formally, the conditions on the spectrum can be expressed as follows.

Theorem 3: A non-trivial set of unitaries \mathbb{U}_C that can be characterized efficiently both in terms of the average number of experiments and the classical computational resources for any number of qutrits exists if the single-qutrit operator basis is maximally partitioning and unitary.

This theorem can be proven straightforwardly from the previous discussion: Since the maximally partitioning unitary basis satisfies conditions 1 and 2 of Theorem 2, then the set of transformations \mathbb{U}_C which allows for efficient characterization contains at least $\mathbb{U}_\delta^0(n)$ and therefore is non-trivial. Moreover, due to the unitary spectrum of the basis operators, the operations in $\mathbb{U}_\delta^0(n)$ satisfy Eq. (12) with $\omega = \exp(2i\pi/p)$ and $a \in [0, p - 1]$. This leads to a generalized uniform relevance distribution, Eq. (22), in all protocols. We show below in Sec. IV that such a generalized uniform relevance distribution yields $\mathcal{C}_{\text{sampl}} \propto \mathcal{O}(1)$.

For a maximally partitioning unitary operator basis, the set of unitaries which leave the basis invariant up to a phase factor is larger than $\mathbb{U}_\delta^0(n)$. This can be inferred from the fact that the operator basis is left invariant, up to a phase factor, also by arbitrary cyclic permutations and those permutations which map basis operators belonging to the same commuting set into each other [29]. Most likely, the set of unitaries given by $\mathbb{U}_\delta^0(n)$ extended by those permutations is also maximal. However, whether this is indeed the case and whether the set coincides with the full group $\mathbb{U}_\delta^\Pi(n)$ of transformations which leaves the set $\mathcal{M}(n)$ of MUB invariant remains an

open question.

F. A unitary operator basis vs actual measurements: The generalized Pauli basis

A maximally partitioning unitary operator basis is the so-called generalized Pauli basis [11, 25, 26]. This basis generates a group under matrix multiplication, the generalized Pauli group. The group of transformations, $\mathbb{U}_C(n)$, leaving the operator basis invariant up to a phase factor, can be identified with the normalizer of the generalized Pauli group, i.e., with the generalized Clifford group [11]. To construct the generalized Pauli basis, one generalizes the standard Pauli σ_z and σ_x operators [11, 25, 26],

$$\begin{aligned} Z(1) &= \omega^n |n\rangle\langle n|, \\ X(1) &= |n+1\rangle\langle n|, \end{aligned} \quad (23)$$

where addition is modulo p , $n \in [0, p - 1]$, and $\omega = \exp(2i\pi/p)$. The generalized Pauli operator basis for a single qutrit is obtained as [30]

$$X^a(1)Z^b(1) \quad a, b = 0, \dots, p - 1. \quad (24)$$

For example, by setting $Y(1) = X(1)Z(1)$ and $V(1) = X(1)Z(1)^2$ the full operator basis for a single qutrit reads $\bar{\mathcal{P}}(1) = \{I(1), X(1), Y(1), Z(1), V(1), X^2(1), Y^2(1), Z^2(1), V^2(1)\}$. Each operator from the set commutes only with itself, its square (corresponding to both its Hermitian conjugate and its inverse) and the identity, i.e., with the operators obtained from a special set of permutations identified in Ref. [29]. This defines for qutrits a unique partitioning into $d + 1 = 4$ sets of commuting operators. The generalized Pauli basis, Eq. (24), gives rise to the definition of the generalized single-qutrit Pauli group as [11, 31]

$$\mathcal{P}(1) = \{\omega^i X^a(1)Z^b(1) \quad a, b, i \in [0, p - 1]\}. \quad (25)$$

In analogy to the qubit case, the Pauli measurements on n qutrits are given by tensor products of the single-qutrit operators, Eq. (24), which are also the generators of the n -qutrit Pauli group.

To summarize, by enforcing unitarity on the λ -matrix in Eq. (21), we can obtain an operator basis which generalizes all the fundamental properties of the standard Pauli operators besides Hermiticity. That is, an orthonormal basis of unitary operators with a maximal partitioning into $d + 1$ commuting subsets which is preserved under tensor product. The p -nary spectrum of the basis is preserved as the number of particles increases, and the operator basis generates a group under matrix multiplication. Since we can define a generalized Clifford group and obtain a uniform relevance distribution, the fundamental requirements for achieving efficient characterization for certain unitaries are met. There are two caveats, however: (i) The Monte Carlo procedure needs

to be generalized for measurement operators with complex eigenvalues. This is done in Appendix B. (ii) Observables have to be Hermitian, so we need to clarify how a unitary, non-Hermitian measurement basis can be connected to measurable observables. There are two options – one can construct Hermitian counterparts of unitary basis operators or utilize the concept of a quantum circuit to simulate a Hermitian measurement.

A Hermitian counterpart can be constructed from the unitary orthonormal set of generalized Pauli operators $\bar{\mathcal{P}}(1) = \{U_k(1)\}_{k=1}^{p^2}$ by noting that for each $U_k(1) \in \bar{\mathcal{P}}(1)$ also $U_k^\dagger(1) = [U_k(1)]^{p-1} \in \bar{\mathcal{P}}(1)$ is contained in $\bar{\mathcal{P}}(1)$. Consequently, a Hermitian orthonormal basis is obtained via the transformation [31]

$$\begin{aligned} H(1) &= (U(1) - U(1)^\dagger)/\sqrt{2}i \\ \bar{H}(1) &= (U(1) + U(1)^\dagger)/\sqrt{2}. \end{aligned} \quad (26)$$

The operators of kind H have spectrum $\text{Im}(\omega^a)$ with $a \in [0, p-1]$, whereas those of kind \bar{H} have spectrum $\text{Re}(\omega^a)$ with $a \in [0, p-1]$. Since $[H(1), U(1)] = [\bar{H}(1), U(1)] = 0$, the partitioning structure of the generalized Pauli basis, and hence the corresponding structure of MUB, is preserved by the transformation (26). However, since Hermiticity is not enforced at the level of the λ matrix, the Hermitian counterpart of the generalized Pauli basis does not inherit the tensor product structure,

$$\begin{aligned} H(n) &= \bigotimes_{i=1}^n U_i(1) = \left(\bigotimes_{i=1}^n U_i(n) - \bigotimes_{i=1}^n U_i^\dagger(n) \right) / i\sqrt{2} \\ &\neq \bigotimes_{i=1}^n H(U_i). \end{aligned} \quad (27)$$

If on one hand this implies that the spectrum of the Hermitian operators remains invariant with respect to the number of qupits on the other the operator basis includes non-local measurements. It is easily seen that, regardless of the number of particles n , the action of a Clifford operation C on the Hermitian basis is $CH(U_k)C^\dagger = H(CU_kC^\dagger)$, since C maps U_k into $CU_kC^\dagger = \omega^i U_{k'}$ with $i \in [0, p-1]$ and U_k^\dagger in $(\omega^i)^* U_{k'}^\dagger$.

In conclusion, a unitary generalization of the Pauli operators maintains all relevant properties of the standard Pauli basis. Despite losing Hermiticity, it can be employed to construct a Hermitian operator basis which, however, does not obey a tensor product structure and hence does not correspond to local measurements. This sets the stage for efficient characterization of qupit Clifford operations. If one uses the unitary generalized Pauli basis, despite the fact that the operators are non-Hermitian, actual measurements can be carried out utilizing the concept of universal quantum circuits [24]: Any measurement of a generalized (non-Hermitian) Pauli operator can be implemented by applying suitable unitary gates to the system coupled to an auxiliary qudit and performing a projective measurement on the auxiliary qudit in the standard basis. The idea of mapping complex spectra to

real measurement results by an appropriate experimental protocol has first been discussed for polarization-path qudits with $p = 4$ [32]. Alternatively to unitary generalized Pauli measurements, the Hermitized version of the basis, Eq. (26), can be adopted. It includes, however, non-local measurements.

IV. EFFICIENT CHARACTERIZATION OF QUDIT OPERATIONS

A. Modifications of the Monte Carlo approach allowing for efficient characterization of qudit operations

When replacing qubits by qudits, only unitary, maximally partitioning operator bases such as the generalized Pauli basis and their Hermitized versions allow for efficient characterization both in terms of $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ and $\mathcal{C}_{\text{sample}} \propto \mathcal{O}(1)$. If a unitary operator basis is chosen, a uniform relevance distribution can be obtained, yielding efficient sampling, by employing a complex generalization of the standard Monte Carlo approach [2, 3, 9]. It is presented in Appendix B.

For a Hermitized basis, the standard Monte Carlo approach needs to be modified at the level of the sampling step. With the standard sampling procedure, efficient sampling cannot be achieved since the relevance distribution of Clifford unitaries in the Hermitized basis is no longer uniform due to the loss of the tensor product structure. We denote the Hermitized basis by $\mathbb{H} = \{\tilde{H}_i\}_{i=1}^{d^2}$ where the \tilde{H}_i comprise both H_i and its Hermitian partner \bar{H}_i . For Clifford operations, the relevance distribution in the Hermitized basis takes on the values

$$P^j(i, k) = \{\text{Re}^2(\omega^a), \text{Im}^2(\omega^a); a \in [0, p-1]\}. \quad (28)$$

For each input operator \tilde{H}_i there are two possible output operators $\tilde{H}_k, \bar{\tilde{H}}_{\bar{k}}$ leading to non-vanishing expectation values. The following relation holds

$$P^j(i, k) + P^j(i, \bar{k}) = 1. \quad (29)$$

It allows for uniform sampling over pairs k, \bar{k} , i.e., one draws uniformly at random an index $i \in [1, d^2]$, selecting the input operator from the set \mathbb{H} . Using a generalization of the Gottesman-Knill theorem [24], one can efficiently compute $C\tilde{H}_iC^\dagger$ where C is the Clifford operation that shall be certified. One thus obtains the indices k, \bar{k} corresponding to the measurements with non-vanishing expectation values and the phase factor ω^a needed to determine the corresponding value of the relevance distribution. At this point, a second sampling step according to Table I is necessary to select a single measurement out of \tilde{H}_k and $\bar{\tilde{H}}_{\bar{k}}$. Such a two-stage sampling is independent of system size. Thus, also for a Hermitized basis, the sampling complexity is $\mathcal{C}_{\text{class}} \propto \mathcal{O}(1)$ and the classical computational resources scale polynomially in n .

| | | |
|--------------------------------------|------------------------------|--------------------------------------|
| | $\tilde{H}_i \in \mathbb{H}$ | $\tilde{H}_i \in \tilde{\mathbb{H}}$ |
| $\tilde{H}_k \in \mathbb{H}$ | $\text{Re}^2(\omega^a)$ | $\text{Im}^2(\omega^a)$ |
| $\tilde{H}_k \in \tilde{\mathbb{H}}$ | $\text{Im}^2(\omega^a)$ | $\text{Re}^2(\omega^a)$ |

TABLE I: Relevance distribution for the additional binary sampling required for the Hermitized version of the unitary operator basis. The symbols \mathbb{H} and $\tilde{\mathbb{H}}$ denote, respectively, the sets of operators of the kind H and \tilde{H} .

B. Hierarchy of operator bases

Our discussion in Section III does not only provide efficient Monte Carlo protocols for the characterization of qudit operations, it also allows to classify all operator bases according to which properties of the standard Pauli basis for qubits they retain. The hierarchy is summarized in Table II.

At the bottom of the hierarchy we find operator bases that only retain Hermiticity, such as the Gell-Mann basis for qutrits. Following Theorem 1, these bases do not allow for efficient Monte Carlo characterization for any unitary. Moreover, they cannot be used in combination with the input-state based protocols that yield a reduction of resources for general unitaries [9]. This follows from the fact that these bases do not allow for the existence of mutually unbiased eigenbases.

The next step in the hierarchy is occupied by Hermitian bases that obey the conditions of Theorem 1. These bases allow for the existence of a set of non-entangling unitaries that can be characterized with $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ in the protocol based on the entanglement fidelity and the one using two classical fidelities. In other words, Theorem 1 ensures that the operator basis admits the existence of non-entangled generalized stabilizer states. This explains why the protocol based on a state 2-design which includes entangled stabilizer states cannot be applied. However, the unitaries for which $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ cannot be characterized efficiently since in general their relevance distribution is not known a priori. Therefore, Monte Carlo characterization with such operator bases still requires classical computational resources that scale exponentially in the number of qudits.

Next, we have Hermitian operator bases which obey the conditions of Theorem 2. These bases enlarge the class of unitaries for which $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ to comprise also entangling operations. They also ensure that this scaling is achieved in all protocols. In other words, enforcing the maximally partitioning property and the condition that all λ must be equal for every n guarantees the existence of both separable and entangled stabilizer states. However, most likely, a Hermitian basis for multiple qudits which is maximally partitioning includes non-local measurements. This would imply that there is no local Hermitian measurement basis allowing to achieve $\langle N_{exp} \rangle \propto \mathcal{O}(1)$ in all protocols. Moreover, even if such a basis existed, it would not allow for efficient characterization of any unitary in terms of the sampling complexity

since the relevance distribution would not be known a priori.

Finally, on top of the hierarchy, we find unitary bases that give rise to a maximal partitioning. These bases retain all the relevant properties of the standard Pauli basis for qubits besides Hermiticity. They allow for efficient characterization in all protocols, provided one generalizes the Monte Carlo procedure to operators with complex eigenvalues. The corresponding class of unitaries comprises not only the elements of $\mathbb{U}_o^0(n)$, mapping elements of two bases into each other, but also certain, if not all, permutations. Efficient Monte Carlo characterization is also achieved by a Hermitized version of such a unitary basis by modifying the sampling to consist of two stages as explained above. The Hermitized version, however, comprises non-local measurements. For generic unitaries, Monte Carlo characterization using Hermitized operator bases requires more computational resources compared to the unitary counterpart. This is due to the loss of the tensor product structure because of which the method of the conditional probabilities [3] can not be applied.

V. CONCLUSIONS

We have shown that there exists a class of unitary operations for multi-level information carriers for which in principle the average fidelity can be estimated efficiently, i.e., with an effort that scales at most polynomially in the number of qudits. However, if the class of unitaries is to comprise entangling operations, the operator basis that must be chosen to allow for efficient characterization is either unitary non-Hermitian or Hermitian but comprising non-local measurements.

Unitary non-Hermitian measurements can be realized via quantum circuits [24, 32]. The corresponding Monte Carlo sampling procedure that is required to carry out the characterization needs to be adapted to complex eigenvalues in the relevance distribution. We have shown that this is straightforward. Employing non-local Hermitian measurements that are constructed out of the unitary operator basis also requires a small modification of the standard Monte Carlo procedure in that a two-stage sampling becomes necessary to achieve a sampling complexity that is independent of system size. Which of the two approaches, unitary circuit measurements or non-local Hermitian measurements, is more practical in an actual experiment remains to be seen.

The crucial feature of operator bases to allow for efficient device characterization is that they give rise to a maximal partitioning of the operators into commuting sets. Fulfilling this condition at the level of single-qudit operators guarantees the existence of a class of unitary transformations that can be characterized with reduced resources in the Monte Carlo protocols based on the entanglement fidelity [2, 3] and two classical fidelities [9]. In that case, a Hermitian basis of local measurements can be utilized. However, in order to achieve efficient

| operator basis | $\langle N_{exp} \rangle$ | \mathcal{C}_{sampl} | local measurements | protocols |
|----------------|---------------------------|---------------------------------------|--------------------|-----------|
| A | $\mathcal{O}(d^2)$ | $\mathcal{O}(n^2 d^4)$ | yes | 1 |
| B | $\mathcal{O}(1)$ | as for general unitaries ^a | yes | 1,2 |
| C | $\mathcal{O}(1)$ | as for general unitaries ^b | most likely not | 1,2,3 |
| D | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | yes | 1,2,3 |
| E | $\mathcal{O}(1)$ | $\mathcal{O}(1)$ | no | 1,2,3 |

^aThe scaling for general unitaries depends on the protocol, cf. Ref. [9].

^bIf a Hermitian basis comprises non-local measurements, then the sampling complexity for general unitaries is increased since the relevance distribution can no longer be computed using conditional probabilities, cf. Ref. [3].

TABLE II: Resources required for characterizing of operations in \mathbb{U}_C . The protocols refer to 1: protocol based on the entanglement fidelity [2, 3], 2: protocol employing two classical fidelities [9], 3: protocol based on a state 2-design [9]. The operator bases are labeled as follows: A: Hermitian bases, such as the Gell-Mann basis for qutrits; B: Hermitian bases constructed as tensor products of a single-qudit bases that give rise to a maximal partitioning with all λ^j in Eq. (21) being equal; C: Hermitian bases that give rise to a maximal partitioning and have equal λ^j for all n ; D: unitary bases that give rise to maximal partitioning and have equal λ^j for all n , such as the generalized Pauli basis; E: Hermitized version of D.

characterization for a larger set of unitaries including entangling operations, the maximally partitioning property needs to be fulfilled at the level of the multi-qudit operators. While it is automatically satisfied by a unitary basis built as tensor product of single-qudit operators that give rise to a maximal partitioning, the same does not appear to be true for Hermitian bases. For the latter, non-local measurements seem unavoidable for efficient characterization of qudit operations.

Our work highlights the intimate relation between the existence of unitaries that can be characterized efficiently and the existence of mutually unbiased bases. In fact, for prime Hilbert space dimensions, that is, at the single qudit level, one can determine a maximal number of such unitaries in a constructive proof [29]. Moreover, our results suggest that the conditions presented in Theorems 1 to 3 are not only sufficient for efficient characterization but also necessary. One might argue that necessity of the maximally partitioning property is questioned by recent results on generalized Pauli bases [24]. Indeed, a generalized Pauli basis, and hence a generalized Clifford group, can be constructed assuming only an arbitrary tensor product decomposition of the Hilbert space, without the necessity of prime subspace dimensions [24]. Since existence of a maximal number of mutually unbiased bases and hence existence of a maximal partitioning is only guaranteed for prime dimensions, such a generalized Clifford group would not be in correspondence with an underlying maximal partitioning structure already at the level of single qudit operators. We believe however that this apparant contradiction can be resolved by considering the tensor product structure assumed in Ref. [24]. Indeed, the properties of a unitary operator basis that is obtained in terms of tensor products over an arbitrary decomposition of the Hilbert space should be equivalent to the properties of the same unitary basis obtained as tensor products over the irreducible decomposition given by the prime factorization. In the irreducible decompo-

sition, each single-qudit generalized Pauli basis gives rise to a maximal partitioning and thus allows for the existence of stabilizer states. This would be consistent with an extension of our theorems in terms of *necessary* conditions for efficient characterization. A rigorous proof of the fact that necessity of the maximal partitioning is consistent with the results of Ref. [24] is beyond the scope of our current work.

Acknowledgments

GG acknowledges financial support from a MIUR-PRIN grant (2010LLKJBX). QSTAR is the MPQ, LENS, IIT, UniFi Joint Center for Quantum Science and Technology in Arcetri.

Appendix A: Proofs

1. Proof of proposition 1

The general form of a unitary transformation between two bases, $A_j, A_{j'} \in \mathcal{M}$ is given by Eq. (19). This expression is general since no ordering of the elements within each basis is specified. What then needs to be proven is that a change of basis between A_j and $A_{j'}$ depends only on the distance between the two indices j and j' , i.e.,

$$U_{jj'} = U_\delta, \quad (\text{A1})$$

where $\delta = j - j'$. This can be done by applying $U_{jj'}$ to a generic element $|\psi_l^i\rangle$ of the basis A_i with $i \in [1, d+1]$ and $l \in [1, d]$,

$$U_{jj'} |\psi_l^i\rangle = \sum_{k=1}^d |\psi_k^{j'}\rangle \langle \psi_k^j | \psi_l^i \rangle. \quad (\text{A2})$$

Without loss of generality [37], one can express $|\psi_l^i\rangle$, using the explicit construction of the mutually unbiased bases for single qubits [26, 27], as follows,

$$|\psi_l^i\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^d (\omega^l)^{d-k} (\omega^{j-i})^{s_k} |\psi_k^j\rangle, \quad (\text{A3})$$

where $\omega = \exp(2i\pi/p)$ and $s_k = \sum_{i=k}^d i$. Equation (A3) implies

$$\langle \psi_k^j | \psi_l^i \rangle = \frac{1}{\sqrt{d}} (\omega^l)^{d-k} (\omega^{j-i})^{s_k}, \quad (\text{A4})$$

which, substituted into Eq. (A2), leads to

$$U_{jj'} |\psi_l^C\rangle = \frac{1}{\sqrt{d}} \sum_k |\psi_k^{j'}\rangle (\omega^l)^{d-k} (\omega^{j-i})^{s_k} = |\psi_l^{i+(j'-j)}\rangle. \quad (\text{A5})$$

Since this argument holds for any $i \in \mathcal{M}$ and any $\delta = j' - j$, one can conclude that indeed Eq. (A1) holds. The same is also true for multiple qudits. This can be shown by using, in Eq. (A3), the general construction of MUB for multiple qubits [27].

Note that, if a precise ordering of the elements within each basis is chosen, the transformation U_δ can be rewritten as

$$U_\delta = \sum_k |\psi_{\Pi(k)}^{j+\delta}\rangle \langle \psi_k^j|, \quad (\text{A6})$$

where $\Pi(k)$ denotes the action of an arbitrary permutation Π on the k th basis index. This yields a decomposition of U_δ in terms of a transformation

$$U_\delta^0 = \sum_k |\psi_k^{j+\delta}\rangle \langle \psi_k^j| \quad (\text{A7})$$

between the k th element of basis j and the k th element of basis $j + \delta$ and permutation Π of the elements of any of the two bases, that is

$$\Pi U_\delta^0 = \sum_{k'k} |\psi_{\pi(k')}^{j+\delta}\rangle \langle \psi_{k'}^{j+\delta} | \psi_k^{j+\delta} \rangle \langle \psi_k^j | = U_\delta. \quad (\text{A8})$$

2. Proof that the unitaries defined by Eq. (19) form a group

The unitaries defined by Eq. (19) form a group, $\mathbb{U}_\delta^\Pi = \{U_\delta\}_{\delta=0}^{d+1}$, under matrix multiplication,

$$U_\delta U_{\delta'} = \sum_{k,l=1}^d |\psi_k^{i+\delta}\rangle \langle \psi_k^i | \psi_l^{i+\delta'} \rangle \langle \psi_l^i | \quad (\text{A9})$$

$$= \sum_{k,l=1}^d |\psi_k^{i+\delta}\rangle \langle \psi_l^i | \frac{(\omega^l)^{d-k}}{\sqrt{d}} (\omega^{-\delta'})^{s_k} \quad (\text{A10})$$

$$= \sum_l |\psi^{i+(\delta+\delta')}\rangle \langle \psi_l^i | = U_{\delta+\delta'}, \quad (\text{A11})$$

where we have used that

$$|\psi^{i+(\delta+\delta')}\rangle = \frac{1}{\sqrt{d}} \sum_k |\psi_k^{i+\delta}\rangle (\omega^l)^{d-k} (\omega^{-\delta'})^{s_k},$$

and

$$U_\delta U_\delta^\dagger = U_\delta U_{-\delta} = \mathbb{1}.$$

Following the same argument, one can conclude that, for a fixed ordering of the elements within each basis, the transformations U_δ^0 also form a group \mathbb{U}_δ^0 and that the full group \mathbb{U}_δ^Π arises as the composition of \mathbb{U}_δ^0 with the group of permutations.

3. Proof of Theorem 1

Let us apply a unitary $U_\delta^0(1)$, Eq. (A7), to a generic element of the single-qubit operator basis,

$$U_\delta^0(1) B_i^j(1) U_{-\delta}^0(1) = \sum_{k=1}^d \lambda_{ik}^j |\psi_k^{j+\delta}\rangle \langle \psi_k^{j+\delta}| = \tilde{B}_i(1). \quad (\text{A12})$$

By definition, $\tilde{B}_i(1)$ belongs to the operator basis $\mathbb{B}(1)$. It corresponds to the element $B_i^{j+\delta}(1)$, up to a phase factor $e^{i\phi_i}$, if and only if $\lambda^j = e^{i\phi_j} \lambda^{j+\delta}$. Since this must be true for every δ , λ^j must be equal to $e^{i\phi_j} \lambda$ for each $j \in [1, d+1]$. Now since each commuting set contains the identity, i.e., the first row of every λ^j is made up of ones, $e^{i\phi_j} = 1$ and $\lambda^j = \lambda$ for each $j \in [1, d+1]$. Since the set of unitaries $\mathbb{U}_\delta^0(1)$ forms a group, the condition on all λ^j to be equal ensures the existence of a group of transformations which leaves the single-qubit operator basis invariant, i.e., $\mathbb{U}_\delta^0(1) \subseteq \mathbb{U}_C(1)$.

Now consider the n -qubit operator basis $\mathbb{B}(n)$, built out of tensor products of the operators in $\mathbb{B}(1)$. The n -qubit operators can be written as $B_i(n) = \bigotimes_{l=1}^n B_{i_l}^{j_l}(1)$, where $B_{i_l}^{j_l}(1)$ denotes a generic single-qubit operator acting on the l th qubit. Existence of the group $\mathbb{U}_\delta^0(1)$ implies that $\mathbb{B}(n)$ is left invariant by the set of unitaries $\tilde{\mathbb{U}}_\delta^0 = \{\tilde{U}_\delta^0(n)\}$ that are built as tensor products of the elements in $\mathbb{U}_\delta^0(1)$. This can be seen as follows: For every $B_i(n)$ and $\tilde{U}_\delta^0(n) \in \tilde{\mathbb{U}}_\delta^0$, one has

$$\begin{aligned} \tilde{U}_\delta^0(n) B_i(n) \tilde{U}_\delta^{0,\dagger}(n) &= \left(\bigotimes_{l=1}^n U_{\delta_l}^0(1) \right) B_i(n) \left(\bigotimes_{l=1}^n U_{\delta_l}^0(1) \right)^\dagger \\ &= \left(\bigotimes_{l=1}^n U_{\delta_l}^0(1) \right) \bigotimes_{l=1}^n B_{i_l}^{j_l}(1) \left(\bigotimes_{l=1}^n U_{-\delta_l}^0(1) \right) \\ &= \bigotimes_{l=1}^n \left(U_{\delta_l}^0(1) B_{i_l}^{j_l}(1) U_{-\delta_l}^0(1) \right) \\ &= \bigotimes_{l=1}^n B_{i_l'}^{j_l'}(1) = B_{i'}(n) \in \mathbb{B}(n). \end{aligned} \quad (\text{A13})$$

This allows to conclude that $\tilde{\mathbb{U}}_\delta^0(n) \subseteq \mathbb{U}_C(n)$. The characteristic function of a generic $\tilde{U}_\delta^0(n) \in \tilde{\mathbb{U}}_\delta^0(n)$ is given by

$$\begin{aligned}\chi_{\tilde{U}_\delta^0(n)}(i, k) &= \frac{1}{d} \text{Tr}[B_k(n) \tilde{U}_\delta^0(n) B_i(n) \tilde{U}_\delta^{0,\dagger}(n)] \\ &= \frac{1}{d} \text{Tr}[B_k(n) B_{i'}(n)] = \delta_{ki'}. \quad (\text{A14})\end{aligned}$$

Therefore these unitaries will lead to a relevance distribution with $d^2 = \mathcal{N}$ non-zero elements in the protocol based on the entanglement fidelity, i.e., formally using input operators [2, 3]. With Eq. (17), one then finds $\langle N_{exp} \rangle \propto \mathcal{O}(1)$. In addition, $\tilde{\mathbb{U}}_\delta^0(n)$ is itself a group since its elements are tensor products of the elements of $\mathbb{U}_\delta^0(1)$.

Let us now check the scaling of the transformations in $\tilde{\mathbb{U}}_\delta^0(n)$ for input-state based protocols. By construction, the operator basis $\mathbb{B}(n)$ admits the existence of $p + 1$ separable mutually unbiased joint eigenbases obtained as tensor products of the elements of the single-qubit bases in $\mathcal{M}(1)$. These $p + 1$ MUB form a subset $\mathcal{M}_{sep}(n)$ of the full set $\mathcal{M}(n)$. By construction, $\mathcal{M}_{sep}(n)$ is mapped into itself by the group of transformations $\tilde{\mathbb{U}}_\delta^0(n)$. Now consider a generic element $|\psi_i^j\rangle$ of a separable basis A_j in $\mathcal{M}_{sep}(n)$. By denoting by $|\psi_{i_l}^{j_l}\rangle$ an element of the joint eigenbasis of the commuting set \mathcal{W}_{j_l} of single-qubit operators acting on the l th qubit, $|\psi_i^j\rangle$ can be expressed as $|\psi_i^j\rangle = \otimes_{l=1}^n |\psi_{i_l}^{j_l}\rangle$. For each state in $\mathcal{M}^{sep}(n)$, the characteristic function of a unitary transformation $\tilde{U}_\delta^0(n) \in \tilde{\mathbb{U}}_\delta^0(n)$ is then

$$\begin{aligned}& \text{Tr} \left[B_i(n) \tilde{U}_\delta^0(n) |\psi_i^j\rangle \langle \psi_k^j | \tilde{U}_\delta^{0,\dagger}(n) \right] \\ &= \text{Tr} \left[B_i(n) |\psi_k^{j'}\rangle \langle \psi_k^{j'} | \right] \\ &= \Pi_{l=1}^n \text{Tr} \left[B_{i_l}^{j_l'}(1) |\psi_{k_l}^{j_l'}\rangle \langle \psi_{k_l}^{j_l'} | \right] \\ &= \begin{cases} E_i(n) & \text{if } j_l' = j_l \forall l \in [1, n], \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A15})\end{aligned}$$

Here $E_i(n) = \Pi_{l=1}^n \lambda_{i_l, k_l}^{j_l'}$ is the eigenvalue of $B_i(n)$ corresponding to the element $|\psi_i^{j'}\rangle$ of the basis $A_{j'}$ in $\mathcal{M}^{sep}(n)$. Provided that the characterization protocol does not require more than $p + 1$ MUB, Eq. (A15) implies that the unitaries in $\tilde{\mathbb{U}}_\delta^0(n)$ correspond to a relevance distribution with $\mathcal{N} = Td$ non-zero elements hence yielding $\langle N_{exp} \rangle \propto \mathcal{O}(1)$. This is the case of the protocol based on classical fidelities since it requires input states from two MUB but not of the 2-design protocol which instead requires the existence of the full set $\mathcal{M}(n)$.

In conclusion, we have proven that, if the maximally partitioning property and the condition that all λ^j 's must be equal are enforced on the single-qubit operator basis, then the existence for any number of qubit of a non-trivial group of unitaries leading to $\langle N_{exp} \rangle \propto \mathcal{O}(1)$, at least in some protocols, is ensured.

Appendix B: Complex Monte Carlo estimation

We abbreviate the values of the characteristic functions, Eq. (6), by

$$\begin{aligned}\alpha_{ik} &= \frac{1}{d} \text{Tr} \left[\mathcal{D}(W_i)^\dagger W_k \right] = \chi_{\mathcal{D}}(i, k) \\ \beta_{ik} &= \frac{1}{d} \text{Tr} \left[U W_i^\dagger U^\dagger W_k \right] = \chi_U(i, k).\end{aligned}$$

In general, α_{ik} and β_{ik} are complex; they are real only if W_k is Hermitian. The average gate fidelity can be expressed in terms of α_{ik} and β_{ik} ,

$$\begin{aligned}F_{av} &= \frac{1}{d^2} \sum_{i,k} \alpha_{ik} \beta_{ik}^* = \sum_{i,k} \frac{|\beta_{ik}|^2}{d^2} \frac{\alpha_{ik}}{\beta_{ik}} \\ &= \sum_{i,k} \text{Pr}(i, k) \frac{\alpha_{ik}}{\beta_{ik}}\end{aligned}$$

with the real-valued relevance distribution

$$\text{Pr}(i, k) = \frac{|\beta_{ik}|^2}{d^2}.$$

Note that if U_0 is a Clifford gate, then for any i there is only a single k such that $\beta_{ik} \neq 0$, taking the value $\frac{1}{d^2}$. For Monte Carlo sampling we define now the complex random variable X on the event space given by the set of tuples (i, k)

$$X(i, k) = \frac{\alpha_{ik}}{\beta_{ik}}. \quad (\text{B1})$$

It is easy to see that the expectation value of this random variable corresponds to F_{av} ,

$$\mathbb{E}(X(i, k)) = \sum_{i,k} \text{Pr}(i, k) \frac{\alpha_{ik}}{\beta_{ik}} = F_{av}. \quad (\text{B2})$$

The Monte Carlo approach seeks an estimate of F_{av} with additive error ϵ and failure probability δ . In other words, one wants to find an estimator Y such that the likelihood that this estimator Y is greater or equal ϵ away from the fidelity F_{av} to be less or equal δ ,

$$\text{Pr}[|Y - F_{av}| \geq \epsilon] \leq \delta. \quad (\text{B3})$$

The complex version of Chebyshev's inequality [33] states that, $\forall t > 0$ and each complex random variable Z with expectation value μ , the following relation is fulfilled

$$\text{Pr}[|Z - \mu| \geq t|\mu|] \leq \frac{\mathbb{E}(ZZ^*) - \mathbb{E}(Z)\mathbb{E}(Z^*)}{t^2|\mu|^2}. \quad (\text{B4})$$

Mapping $t > 0$ onto $t|\mu| \equiv \kappa > 0$ leads to

$$\text{Pr}[|Z - \mu| \geq \kappa] \leq \frac{\mathbb{E}(ZZ^*) - \mathbb{E}(Z)\mathbb{E}(Z^*)}{\kappa^2}. \quad (\text{B5})$$

Now one just needs to find a suitable estimator Y and calculate its expectation value and variance.

To this end, set the number of draws L from the event space given by the tuples (i, k) to $L = \lceil \frac{1}{\epsilon^2 \delta} \rceil$ where $\lceil \cdot \rceil$ means to round up to the nearest integer. Choosing independently some events $(i_1, k_1), \dots, (i_L, k_L)$ out of the total event space yields independent estimates $X_1 = \frac{\alpha_{i_1 k_1}}{\beta_{i_1 k_1}}, \dots, X_L = \frac{\alpha_{i_L k_L}}{\beta_{i_L k_L}}$. Now define $Y = \frac{1}{L} \sum_{l=1}^L X_l$. We explain how to estimate Y which in turn is an approximation to F_{av} . Note that Y structurally resembles X . However, the relevance distribution does not appear. This is due to the fact that each X_l is already chosen with the corresponding probability. Hence in the limit of $L \rightarrow \infty$: $Y \rightarrow X$.

Consider the choice of (i_l, k_l) with $l = 1, \dots, L$ chosen as explained above and i_l denoting the index of the input operator of the l th measurement by k_l the index of the measured operator of the l th measurement. For each l the operator W_{k_l} will be measured on the state that is obtained by sending a randomly drawn eigenstate $|\phi_a^{i_l}\rangle$ of W_{i_l} with corresponding eigenvalue $\lambda_a^{i_l}$ through the device (a is drawn out of the set $\{1, \dots, d\}$). This is repeated a total number of m_l times where

$$m_l = \left\lceil \frac{4}{|\beta_{i_l k_l}|^2 L \epsilon^2} \log \left(\frac{4}{\delta} \right) \right\rceil. \quad (\text{B6})$$

This choice of m_l guarantees that Eq. (B3) is fulfilled as we show below. Note that each measurement gives an eigenvalue of the operator W_{k_l} . We denote these, in general complex, measurement results by w_{ln} with n referring to the n th repetition of the l th measurement. Each of these measurements results in an eigenvalue $w_{ln} \in \text{spec}(W_{k_l})$. We assume the expectation value of a measurement of an operator W_{k_l} for a state ρ to be given by

$$\langle W_{k_l} \rangle_\rho = \text{Tr} [\rho^\dagger W_{k_l}] = \text{Tr} [\rho W_{k_l}]$$

also for non-Hermitian operators. Let us define now $A_{ln} = (\lambda_{a_n}^{i_l})^* w_{ln}$ where $\lambda_{a_n}^{i_l}$ is the eigenvalue corresponding to the eigenstate $|\phi_{a_n}^{i_l}\rangle$ of the operator W_{i_l} . Note that

$$\begin{aligned} \mathbb{E}(A_{ln}) &= \frac{1}{d} \sum_{a_n=1}^d (\lambda_{a_n}^{i_l})^* w_{ln} \\ &= \frac{1}{d} \sum_{a_n=1}^d (\lambda_{a_n}^{i_l})^* \text{Tr} [\mathcal{D}(|\phi_{a_n}^{i_l}\rangle \langle \phi_{a_n}^{i_l}|)^\dagger W_{k_l}] \\ &= \frac{1}{d} \sum_{a_n=1}^d \text{Tr} [(\lambda_{a_n}^{i_l})^* \mathcal{D}(|\phi_{a_n}^{i_l}\rangle \langle \phi_{a_n}^{i_l}|)^\dagger W_{k_l}] \\ &= \frac{1}{d} \text{Tr} \left[\mathcal{D} \left(\sum_{a_n=1}^d \lambda_{a_n}^{i_l} |\phi_{a_n}^{i_l}\rangle \langle \phi_{a_n}^{i_l}| \right)^\dagger W_{k_l} \right] \\ &= \frac{1}{d} \text{Tr} [\mathcal{D}(W_{i_l})^\dagger W_{k_l}] = \alpha_{i_l k_l}. \end{aligned}$$

An approximation to X_l , denoted by \tilde{X}_l , can now be

introduced,

$$\tilde{X}_l = \frac{1}{\beta_{i_l k_l}} \cdot \frac{1}{m_l} \sum_{n=1}^{m_l} A_{ln}. \quad (\text{B7})$$

Since $\mathbb{E}(B_{ln}) \equiv \langle A_{ln} \rangle = \alpha_{i_l k_l}$, it is clear that $\frac{1}{m_l} \sum_{n=1}^{m_l} A_{ln} \rightarrow \alpha_{i_l k_l}$.

For the final step in the Monte Carlo estimation, let

$$\tilde{Y} = \frac{1}{L} \sum_{l=1}^L \tilde{X}_l. \quad (\text{B8})$$

Just like \tilde{X}_l is an approximation to X_l , \tilde{Y} is an approximation to Y or in other words an estimate for Y . The goal is to fulfill Hoeffding's inequality, which we prove below,

$$\Pr \left[|\tilde{Y} - Y| \geq \epsilon \right] \leq \delta. \quad (\text{B9})$$

The whole procedure uses the channel a total number of $m = \sum_{l=1}^L m_l$ times. This value in estimation can be bounded by calculating $\mathbb{E}(m_l)$ which is the expected number of experimental repetitions for the given setting (i_l, k_l) . In other words $\mathbb{E}(m_l)$ is the number of experiments one has to perform for a setting (i, k) multiplied by the probability that this setting is chosen. Denoting by $m_l(i, k)$ the number of experiments for the tuple (i, k) , given by Eq. (B6), the expectation value becomes

$$\begin{aligned} \mathbb{E}(m_l) &= \sum_{ik} \Pr(i, k) m_l(i, k) \\ &= \frac{1}{d^2} \sum_{ik} |\beta_{ik}|^2 \left\lceil \frac{4}{|\beta_{ik}|^2 L \epsilon^2} \log \left(\frac{4}{\delta} \right) \right\rceil \quad (\text{B10}) \\ &\leq 1 + \frac{4d^2}{L\epsilon^2} \log \left(\frac{4}{\delta} \right), \end{aligned}$$

where 1 accounts for the fact that the smallest integer greater than the expression in the brackets $\lceil \cdot \rceil$ is taken. The total number of experiments given by the sum of all m_l is found to be

$$\begin{aligned} \mathbb{E}(m) &= \sum_{l=1}^L \mathbb{E}(m_l) \leq L \cdot \left[1 + \frac{4d^2}{L\epsilon^2} \log \left(\frac{4}{\delta} \right) \right] \\ &\leq 1 + \frac{1}{\epsilon^2 \delta} + \frac{4d^2}{\epsilon^2} \log \left(\frac{4}{\delta} \right), \quad (\text{B11}) \end{aligned}$$

where 1 appears for the same reason as above. Note that this scales as $\mathcal{O}(d^2)$. For Clifford gates, there are only d^2 nonvanishing entries the sum in Eq. (B10) since for each k there exists only one l for which $\beta_{kl} \neq 0$. This leads to

$$\mathbb{E}(m_l) \leq 1 + \frac{4}{L\epsilon^2} \log \left(\frac{4}{\delta} \right)$$

and consequently

$$\mathbb{E}(m) \leq 1 + \frac{1}{\epsilon^2 \delta} + \frac{4}{\epsilon^2} \log \left(\frac{4}{\delta} \right),$$

resulting in a scaling of $\mathcal{O}(1)$.

Finally we prove validity of Eqs. (B3) and (B9). We first consider Eq. (B3), where the numerator of the right

hand side of the Chebyshev inequality needs to be estimated for $Z = X_l$,

$$\begin{aligned} \mathbb{E}(X_l X_l^*) - \mathbb{E}(X_l) \mathbb{E}(X_l^*) &= \sum_{ik} \Pr(i, k) \frac{|\alpha_{ik}|^2}{|\beta_{ik}|^2} - \left| \sum_{ik} \Pr(i, k) \frac{\alpha_{ik}}{\beta_{ik}} \right|^2 = \frac{1}{d^2} \sum_{ik} |\alpha_{ik}|^2 - F^2 \\ &= \frac{1}{d^4} \sum_{ik} \langle \mathcal{D}(W_i) \| W_k \rangle \langle W_k \| \mathcal{D}(W_i) \rangle - F_H^2 = \frac{1}{d^4} \sum_{ik} \left| \text{Tr} [W_k^\dagger \mathcal{D}(W_i)] \right|^2 - F^2 \end{aligned}$$

Obviously $0 \leq F \leq 1 \implies 0 \leq F^2 \leq 1$ for all fidelities discussed in this paper. The same is true for the first term. This can be seen most easily in terms of the process matrix. For any operator O , one can write

$$\mathcal{D}(O) = \sum_{nm} \chi_{nm} W_m O W_n^\dagger.$$

Clearly for $O = W_i$,

$$\mathcal{D}(W_i) = \sum_{nm} \chi_{nm} W_m W_i W_n^\dagger.$$

It follows that

$$\begin{aligned} \left| \text{Tr} [W_k^\dagger \mathcal{D}(W_i)] \right|^2 &= \left| \sum_{nm} \chi_{nm} \text{Tr} [W_k^\dagger W_m W_i W_n^\dagger] \right|^2 \\ &\leq \sum_{nm} |\chi_{nm}|^2 \left| \text{Tr} [W_k^\dagger W_m W_i W_n^\dagger] \right|^2 \end{aligned}$$

For fixed i and k , the operator $W_k^\dagger W_m W_i$ is proportional to a Pauli operator. Consider the expression

$$\sum_{ik} \left| \text{Tr} [W_k^\dagger W_m W_i W_n^\dagger] \right|^2.$$

For fixed m, n and a certain i there exists exactly one k such that this is nonzero, namely if and only if

$$W_k^\dagger W_m W_i W_n^\dagger \sim \mathbb{1}_d. \quad (\text{B12})$$

That is,

$$W_k \sim W_m W_i W_n^\dagger.$$

Due to orthonormality of the operator basis, there is only one such k for which this relation can be fulfilled. For Pauli operators the proportionality constant has modulus 1, hence

$$\sum_{ik} \left| \text{Tr} [W_k^\dagger W_m W_i W_n^\dagger] \right|^2 = d^2 \cdot d^2 = d^4.$$

This results in a trace of d for the d^2 tuple (i, k) for which relation (B12) holds. Consequently

$$\frac{1}{d^4} \sum_{ik} \left| \text{Tr} [W_k^\dagger \mathcal{D}(W_i)] \right|^2 \leq \sum_{ik} |\chi_{ik}|^2.$$

Due the Choi-Jamiolkowsky isomorphism, the process matrix corresponds to a density matrix in the d^2 -dimensional Hilbert space $\mathcal{H} \otimes \mathcal{H}$. It can easily be seen that $\sum_{ik} |\chi_{ik}|^2$ corresponds to the purity of this density matrix which cannot be greater than 1. Therefore

$$\frac{1}{d^4} \sum_{ik} \left| \text{Tr} [W_k^\dagger \mathcal{D}(W_i)] \right|^2 \leq 1.$$

Hence $[\mathbb{E}(X_l X_l^*) - \mathbb{E}(X_l) \mathbb{E}(X_l^*)]$ is the difference between two numbers in the interval $[0, 1]$ and consequently smaller than 1,

$$\mathbb{E}(X_l X_l^*) - \mathbb{E}(X_l) \mathbb{E}(X_l^*) \leq 1.$$

It follows for $Y = Y = \frac{1}{L} \sum_{l=1}^L X_l$ that

$$\begin{aligned}
\mathbb{E}(YY^*) - \mathbb{E}(Y)\mathbb{E}(Y^*) &= \mathbb{E}\left(\left(\frac{1}{L}\sum_l X_l\right)\left(\frac{1}{L}\sum_{l'} X_{l'}^*\right)\right) - \mathbb{E}\left(\frac{1}{L}\sum_l X_l\right)\mathbb{E}\left(\frac{1}{L}\sum_{l'} X_{l'}^*\right) \\
&= \frac{1}{L^2}\sum_{l,l'} \mathbb{E}(X_l X_{l'}^*) - \frac{1}{L^2}\sum_{l,l'} \mathbb{E}(X_l)\mathbb{E}(X_{l'}^*) \\
&= \frac{1}{L^2}\sum_{l,l'} [\mathbb{E}(X_l X_{l'}) - \mathbb{E}(X_l)\mathbb{E}(X_{l'}^*)] = \frac{1}{L^2}\sum_l [\mathbb{E}(X_l X_l) - \mathbb{E}(X_l)\mathbb{E}(X_l^*)] \\
&\leq \frac{L}{L^2} = \frac{1}{L}
\end{aligned}$$

where use has been made of $\mathbb{E}(X_l X_{l'}) = \mathbb{E}(X_l)\mathbb{E}(X_{l'})$ for the $X_l \neq X_{l'}$ which are uncorrelated. Chebyshev's inequality, Eq. (B4), consequently yields

$$\Pr[|Y - F| \geq \kappa] \leq \frac{1}{L\kappa^2}. \quad (\text{B13})$$

Now set $\kappa = \sqrt{\frac{1}{L\delta}}$ and $L = \frac{1}{\epsilon^2\delta}$ to obtain

$$\Pr[|Y - F| \geq \epsilon] \leq \delta$$

To show the validity of Eq. (B9) we use the complex version of Hoeffding's inequality [34].

Lemma: Let $\vec{a} \in \mathbb{R}^n$ and $\{X_i\}_{i=1,\dots,n}$ be independent zero-mean complex-valued random variables with $\forall i: |X_i| \leq a_i$. Then $\forall \delta > 0$

$$\Pr\left(\left|\sum_{i=1}^N X_i\right| \geq \delta\right) \leq 4 \exp\left(-\frac{\delta^2}{4\sum_{i=1}^n |a_i|^2}\right)$$

Corollary: Let $\vec{a} \in \mathbb{R}^n$ and $\{X_i\}_{i=1,\dots,n}$ be independent complex-valued random variables with mean value $\sum_{i=1}^N \langle X_i \rangle = \langle X \rangle$ where $X = \sum_{i=1}^N X_i$ and $\forall i: |X_i - \langle X_i \rangle| \leq a_i$. Then $\forall \delta > 0$

$$\Pr(|X - \langle X \rangle| \geq \delta) \leq 4 \exp\left(-\frac{\delta^2}{4\sum_{i=1}^n |a_i|^2}\right)$$

Proof: Apply Hoeffding's inequality to the random variables $X_i - \langle X_i \rangle$.

Specifically this means for $\delta > 0$, $n = L$ and $\tilde{Y} = \frac{1}{L}\sum_{l=1}^L \tilde{X}_l$ with $\langle \tilde{Y} \rangle = \frac{1}{L}\sum_{l=1}^L \langle \tilde{X}_l \rangle = \frac{1}{L}\sum_{l=1}^L X_l =$

Y . Note furthermore that the \tilde{X}_l are composed as a sum themselves of independent random variables A_{ln} corresponding to measurement results with modulus smaller than 1 and expectation value with modulus smaller than 1. As such we can write

$$\Pr[|\tilde{Y} - Y| \geq \epsilon] \leq 4 \exp\left(-\frac{4\epsilon^2}{C}\right) \quad (\text{B14})$$

where

$$C = \sum_{l=1}^L \frac{1}{L} m_l |2c_l|^2, \quad c_l = \frac{1}{m_l \beta_{l k_l}} \quad (\text{B15})$$

since $[A_{ln} - \langle A_{ln} \rangle]$, as discussed for Eq. (B7), always takes values with modulus smaller than 2.

Calculating C leads to

$$\begin{aligned}
C &= \sum_{l=1}^L \frac{4}{L^2 \beta_{l k_l}^2 m_l} = \sum_{l=1}^L \frac{4\beta_{l k_l}^2 L \epsilon^2}{4L^2 \beta_{l k_l}^2 \log\left(\frac{4}{\delta}\right)} \\
&= \sum_{l=1}^L \frac{\epsilon^2}{L \log\left(\frac{4}{\delta}\right)} = \frac{\epsilon^2}{\log\left(\frac{4}{\delta}\right)}. \quad (\text{B16})
\end{aligned}$$

Plugging this into Hoeffding's inequality yields

$$\begin{aligned}
\Pr[|\tilde{Y} - Y| \geq \epsilon] &\leq 4 \exp\left(-\frac{4\epsilon^2}{C}\right) = 4 \exp\left(-4 \log\left(\frac{4}{\delta}\right)\right) \\
&\leq 4 \exp\left(\log\left(\frac{\delta^2}{16}\right)\right) = \frac{\delta^2}{4} \leq \delta. \quad (\text{B17})
\end{aligned}$$

Hence the failure probability is $\leq \delta$ as desired.

-
- [1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
 - [2] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).
 - [3] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).

- [4] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Phys. Rev. Lett. **106**, 100401 (2011).
- [5] C. T. Schmiegelow, A. Bendersky, M. A. Larotonda, and J. P. Paz, Phys. Rev. Lett. **107**, 100502 (2011).
- [6] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).

- [7] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
- [8] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100**, 190403 (2008).
- [9] D. M. Reich, G. Gualdi, and C. P. Koch, Phys. Rev. Lett. **111**, 200401 (2013).
- [10] H. F. Hofmann, Phys. Rev. Lett. **94**, 160504 (2005).
- [11] D. Gottesman, Chaos, Solitons & Fractals **10**, 1749 (1999).
- [12] P. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan, Information Processing Letters **75**, 101 (2000), ISSN 0020-0190.
- [13] D. Gottesman, in *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, edited by S. P. Corney, R. Delbourgo, and P. D. Jarvis (International Press, Cambridge, MA, 1999), pp. 32–43, arXiv:quant-ph/9807006.
- [14] M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, E. Lucero, A. D. O’Connell, D. Sank, H. Wang, J. Wenner, A. N. Cleland, et al., Science **325**, 722 (2009).
- [15] F. W. Strauch, Phys. Rev. Lett. **109**, 210501 (2012).
- [16] G. Molina-Terriza, A. Vaziri, J. Řeháček, Z. Hradil, and A. Zeilinger, Phys. Rev. Lett. **92**, 167903 (2004).
- [17] S. Gröblacher, T. Jennewein, A. Vaziri, G. Weihs, and A. Zeilinger, New J. Phys. **8**, 75 (2006).
- [18] T. C. Ralph, K. J. Resch, and A. Gilchrist, Phys. Rev. A **75**, 022313 (2007).
- [19] B. P. Lanyon, T. J. Weinhold, N. K. Langford, J. L. O’Brien, K. J. Resch, A. Gilchrist, and A. G. White, Phys. Rev. Lett. **100**, 060504 (2008).
- [20] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).
- [21] M. Nielsen, Physics Letters A **303**, 249 (2002).
- [22] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).
- [23] D. Gottesman, Phys. Rev. A **57**, 127 (1998).
- [24] J. Bermejo-Vega and M. Van Den Nest, Quant. Info. Comput. **14**, 0181 (2014).
- [25] J. Lawrence, C. Brukner, and A. Zeilinger, Phys. Rev. A **65**, 032320 (2002).
- [26] S. Bandyopadhyay, P. O. Boykin, and V. V. F. Roychowdhury, Algorithmica **34**, 512 (2002).
- [27] W. K. Wootters and B. D. Fields, Ann. Phys. **191**, 363 (1989).
- [28] J. Lawrence, Phys. Rev. A **84**, 022338 (2011).
- [29] D. M. Reich, G. Gualdi, and C. P. Koch, arXiv:1403.7154 (2014).
- [30] D. Gottesman, A. Kitaev, and P. J., Phys. Rev. A **64**, 012310 (2001).
- [31] J. Lawrence, Phys. Rev. A **70**, 012302 (2004).
- [32] T. Paterek, Physics Letters A **367**, 57 (2007).
- [33] M. Manjunath, K. Mehlhorn, K. Panagiotou, and H. Sun, in *Proceedings of the 19th Annual European Symposium on Algorithms (ESA)*, edited by C. Demetrescu and M. M. Halldorsson (Springer, 2011), vol. 6942 of *Lecture Notes in Computer Science*, p. 677.
- [34] S. T. Li, S. Oymak, and B. Hassibi, Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing p. 3817 (2012).
- [35] Even though tensor products of single-qubit Pauli operators contain entangling operations, each Pauli operator can be measured in a separable eigenbasis.
- [36] The maximally partitioning property also allows for an explicit construction of the $d + 1$ MUB.
- [37] While, to the best of our knowledge, no strict proof has been reported that all sets of mutually unbiased bases obey the form of Eq. (A3) or an equivalent, we believe that the proof still works in the general case.